A Bayesian-motivated test for linear model in high-dimensional setting

Rui Wang[1] and Xingzhong Xu[*1,2]

[1] School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081,China

[2] Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China

Friday 21st December, 2018

Abstract

Using the idea of Bayesian factor, a new test for linear model in high-dimensional setting is proposed.

Our theory is also useful in.

1 Introduction

Consider the high-dimensional linear regression model of the form

$$\mathbf{y} = \mathbf{X}_a \boldsymbol{\beta}_a + \mathbf{X}_b \boldsymbol{\beta}_b + \boldsymbol{\varepsilon},$$

where $\mathbf{y} \in \mathbb{R}^n$ is the response, $\mathbf{X}_a$, $\mathbf{X}_b$ are $n \times q$ and $n \times p$ design matrices, respectively, $\boldsymbol{\beta}_a \in \mathbb{R}^q$, $\boldsymbol{\beta}_b \in \mathbb{R}^p$ are unknown regression coefficients, and $\boldsymbol{\varepsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$ are the iid errors with mean 0 and covariance $\sigma^2 = \phi^{-1}$. Here we break the predictors into two parts $\mathbf{X}_a$ and $\mathbf{X}_b$ such that $\mathbf{X}_a$ contains the predictors that are known to have effect on the response, and we would like to know if $\mathbf{X}_b$ contains useful predictors. That is, we are interested in testing the hypotheses

$$\mathcal{H}_0 : \boldsymbol{\beta}_b = 0, \quad \text{v.s.} \quad \mathcal{H}_1 : \boldsymbol{\beta}_b \neq 0. \tag{1}$$

Motivated by many recent applications of high dimensional regression, we consider the situation where $p + q$ is much larger than n.

[*] Corresponding author

Email address: xuxz@bit.edu.cn

1.1 Overview of existing tests

The conventional test for hypotheses (1) is the F-test which is also the likelihood ratio test under normality. However, the F-test is not well defined in high dimensional setting. In fact, if $\boldsymbol{\varepsilon}$ is normal distributed and $\text{Rank}[\mathbf{X}_a; \mathbf{X}_b] = n$, then the likelihood is unbounded under the alternative hypothesis. This calls for new test methodologies in high-dimensional setting.

Two different high-dimensional settings have been extensively considered in the literature. One is the small p, large q setting. An important example of this setting is testing individual coefficients of a high-dimensional regression. See Buhlmann (2013), Zhang and Zhang (2014) and Lan et al. (2016) for testing procedures in this setting. In this paper, however, we focus on the other setting, namely the large p, small q setting. In this case, there are just a few covariates, namely $\mathbf{X}_a$, are known to have effect on the response, while there remain a large number of covariates, namely $\mathbf{X}_b$, to be tested. In practice, which covariates belong to the part $\mathbf{X}_a$ is determined apriori. If no prior knowledge is available, $\mathbf{X}_a$ can be $\mathbf{1}_n$.

Many test procedures have been proposed in the large p, small q setting. Based on an empirical Bayes model, Goeman et al. (2006) and Goeman et al. (2011) proposed a score test as well as a method to determine the critical value of their test statistic. Later, Lan et al. (2014) proposed a similar test, but using normal distribution to determine the critical value. There are also many other lines of search. Zhong and Chen (2011) proposed a test based on U-statistics for the case $\mathbf{X}_a = \mathbf{1}_n$. Later, Wang and Cui (2015) generalized their test for general design matrix $\mathbf{X}_a$. Scalar invariant tests: Feng et al. (2013) and Xu (2016). Debiased tests: .

Also, many Bayesian hypothesis testing procedures have been proposed in the literature. See JAVIER GIRÓN et al. (2006); Goddard and Johnson (2016); Zhou and Guan (2018) and the references therein.

1.2 Main contributions

Bayesian methods may be powerful tools in high dimensional setting.

1.3 Organization of the paper

The proposed test is the limit of Bayes factors.

2 Methodology

As Goeman et al. (2006) pointed out, if $\boldsymbol{\beta}_b \neq 0$ but $\mathbf{X}_b\boldsymbol{\beta}_b = 0$, no test has any power. Goeman et al. (2006) used Bayesian method. Their idea is to choose an 'unbiased' distribution of $\boldsymbol{\beta}_b$. As they noticed, their test has negligible power for many alternatives, and is not unbiased. In theory, we prove that there is no nontrivial unbiased test.

The following proposition implies that there is no nontrivial unbiased test.

Proposition 1. *Suppose $\mathbf{y} \sim \mathcal{N}_n(\mu, \phi^{-1}\mathbf{I}_n)$. We test $H_0 : \mu = \mathbf{X}_a\boldsymbol{\beta}_a, \boldsymbol{\beta}_a \in \mathbb{R}^q$ versus $H_1 : \mu \in \mathbb{R}^n$, where $\mathbf{X}_a$ is an $n \times q$ matrix with full column rank, $q < n$. Let $\varphi(\mathbf{y})$ be a test function, that is, a Borel measurable function, $0 \leq \phi(\mathbf{y}) \leq 1$. If $\int \varphi(\mathbf{y})\mathcal{N}_n(\mathbf{X}_a\boldsymbol{\beta}_a, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) = \alpha$ for $\boldsymbol{\beta}_a \in \mathbb{R}^q$, $\phi > 0$ and $\int \varphi(\mathbf{y})\mathcal{N}_n(\mu, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) \geq \alpha$ for $\mu \in \mathbb{R}^n$, $\phi > 0$, then $\varphi(\mathbf{y}) = \alpha$, a.s.*

So we can not find a universally good test. Instead, we would like to find a test with good average behaviour. So Bayesian methods are natural choices in this case.

Bayes hypothesis testing use the Bayes factor.

$$B_{10} = \frac{\int f_1(\mathbf{y}|\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi)\pi_1(\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi)d\boldsymbol{\beta}_b d\boldsymbol{\beta}_a d\phi}{\int f_0(\mathbf{y}|\boldsymbol{\beta}_a, \phi)\pi_0(\boldsymbol{\beta}_a, \phi)d\boldsymbol{\beta}_a d\phi}.$$

There have been several extensions of g-priors to $p > n$ case: Maruyama and George (2011), Shang and Clayton (2011).

Under H_0, we impose the reference prior $\pi_0(\boldsymbol{\beta}_a, \phi) = c/\phi$. Note that under H_1, the posterior corresponding to the referece prior is proper if and only if $\text{Rank}(\mathbf{X}_a, \mathbf{X}_b) = q + p$ and $n > q + p$. That is, the minimal training sample size is $q+p+1$. So we cannot impose the reference prior under H_1 provided $q + p \geq n$. We temporarily impose the conditional prior $\boldsymbol{\beta}_b|\boldsymbol{\beta}_a, \phi \sim \mathcal{N}_p(0, \kappa^{-1}\phi^{-1}\mathbf{I}_p)$. There are extansive literature consider the choice of κ. Kass and Wasserman (1995) choose κ such that the amount of information about the parameter equal to the amount of information contained in one observation. Thus, under H_1, we put the prior

$$\pi_1(\boldsymbol{\beta}_b|\boldsymbol{\beta}_a, \phi) = \mathcal{N}_p\left(0, \frac{1}{\kappa\phi}\mathbf{I}_p\right)(\boldsymbol{\beta}_b), \quad \pi_1(\boldsymbol{\beta}_a, \phi) = \frac{c}{\phi}.$$

$$\begin{aligned} m_0(\mathbf{y}; \kappa, \tau) &:= \int f_0^\tau(\mathbf{y}|\boldsymbol{\beta}_a, \phi)\pi_0(\boldsymbol{\beta}_a, \phi)d\boldsymbol{\beta}_a d\phi \\ &= \frac{c_0\Gamma\left(\frac{\tau n - q}{2}\right)}{\pi^{\frac{\tau n - q}{2}}\tau^{\frac{\tau n}{2}}|\mathbf{X}_a^\top\mathbf{X}_a|^{\frac{1}{2}}\|(\mathbf{I}_n - \mathbf{P}_a)\mathbf{y}\|^{\tau n - q}}. \end{aligned}$$

$$\begin{aligned} m_1(\mathbf{y}; \kappa, \tau) &:= \int f_1^\tau(\mathbf{y}|\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi)\pi_1(\boldsymbol{\beta}_b|\boldsymbol{\beta}_a, \phi)\pi_1(\boldsymbol{\beta}_a, \phi)d\boldsymbol{\beta}_a d\boldsymbol{\beta}_b d\phi \\ &= \frac{c_1\kappa^{\frac{p}{2}}\Gamma\left(\frac{\tau n - q}{2}\right)}{\pi^{\frac{\tau n - q}{2}}\tau^{\frac{\tau n + p}{2}}|\mathbf{X}_a^\top\mathbf{X}_a|^{\frac{1}{2}}|\mathbf{X}_b^{*\top}\mathbf{X}_b^* + \frac{\kappa}{\tau}\mathbf{I}_p|^{\frac{1}{2}}} \frac{1}{\left[\mathbf{y}^{*\top}\mathbf{y}^* - \mathbf{y}^{*\top}\mathbf{X}_b^*(\mathbf{X}_b^{*\top}\mathbf{X}_b^* + \frac{\kappa}{\tau}\mathbf{I}_p)^{-1}\mathbf{X}_b^{*\top}\mathbf{y}^*\right]^{\frac{\tau n - q}{2}}}. \end{aligned}$$

$$\frac{m_1(\mathbf{y}; \kappa, \tau)}{m_0(\mathbf{y}; \kappa, \tau)} = \frac{c_1\kappa^{\frac{p}{2}}}{c_0\tau^{\frac{p}{2}}|\mathbf{X}_b^{*\top}\mathbf{X}_b^* + \frac{\kappa}{\tau}\mathbf{I}_p|^{\frac{1}{2}}}\left(\frac{\mathbf{y}^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^* - \mathbf{y}^{*\top}\mathbf{X}_b^*(\mathbf{X}_b^{*\top}\mathbf{X}_b^* + \frac{\kappa}{\tau}\mathbf{I}_p)^{-1}\mathbf{X}_b^{*\top}\mathbf{y}^*}\right)^{\frac{\tau n - q}{2}}$$

It is straightforward to show that the Bayes factor associated with these priors is

$$\begin{aligned} B_{10}^\kappa = &\frac{\kappa^{p/2}}{|\mathbf{X}_b^\top(\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b + \kappa\mathbf{I}_p|^{1/2}} \cdot \\ &\left(\frac{\mathbf{y}\top(\mathbf{I}_n - \mathbf{P}_a)\mathbf{y}}{\mathbf{y}\top(\mathbf{I}_n - \mathbf{P}_a)\mathbf{y} - \mathbf{y}\top(\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b\left(\mathbf{X}_b^\top(\mathbf{I} - \mathbf{P}_a)\mathbf{X}_b + \kappa\mathbf{I}_p\right)^{-1}\mathbf{X}_b^\top(\mathbf{I}_n - \mathbf{P}_a)\mathbf{y}}\right)^{(n-q)/2}. \end{aligned}$$

Thus,

$$2\log B_{10}^{\kappa} = p\log\kappa - \log|\mathbf{X}_b^\top(\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b + \kappa\mathbf{I}_p|$$
$$- (n-q)\log\left(1 - \frac{\mathbf{y}\top(\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b\left(\mathbf{X}_b^\top(\mathbf{I} - \mathbf{P}_a)\mathbf{X}_b + \kappa\mathbf{I}_p\right)^{-1}\mathbf{X}_b^\top(\mathbf{I}_n - \mathbf{P}_a)\mathbf{y}}{\mathbf{y}\top(\mathbf{I}_n - \mathbf{P}_a)\mathbf{y}}\right).$$

Denote by $\mathbf{I}_n - \mathbf{P}_a = \tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top$ the rank decomposition of $\mathbf{I}_n - \mathbf{P}_a$, where $\tilde{\mathbf{U}}_a$ is a $n \times (n-q)$ column orthogonal matrix. Let $\mathbf{X}_b^* = \tilde{\mathbf{U}}_a^\top\mathbf{X}_b$, $\mathbf{y}^* = \tilde{\mathbf{U}}_a^\top\mathbf{y}$. Let γ_i be the ith largest eigenvalue of $\mathbf{X}_b^*\mathbf{X}_b^{*\top}$, $i = 1, \ldots, n-q$. Denote by $\mathbf{X}_b^* = \mathbf{U}_b^*\mathbf{D}_b^*\mathbf{V}_b^{*\top}$ the singular value decomposition of $\mathbf{X}_b^*$, where $\mathbf{U}_b^*$, $\mathbf{V}_b^*$ are $(n-q)\times(n-q)$ and $p \times (n-q)$ column orthogonal matrices, respectively, and $\mathbf{D}_b^* = \operatorname{diag}(\sqrt{\gamma_1}, \ldots, \sqrt{\gamma_{n-q}})$. Then

$$2\log B_{10}^{\kappa} = p\log\kappa - \sum_{i=1}^{n-q}\log(\gamma_i + \kappa) - (p - (n-q))\log\kappa$$
$$- (n-q)\log\left(1 - \frac{\mathbf{y}^{*\top}\mathbf{X}_b^*\left(\mathbf{X}_b^{*\top}\mathbf{X}_b^* + \kappa\mathbf{I}_p\right)^{-1}\mathbf{X}_b^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}\right)$$
$$= -\sum_{i=1}^{n-q}\log(\gamma_i + \kappa) + (n-q)\log\left(\frac{\mathbf{y}^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{U}_b^*\left[\frac{1}{\kappa}\left(\mathbf{I}_{n-q} - \mathbf{D}_b^*\left(\mathbf{D}_b^{*2} + \kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^*\right)\right]\mathbf{U}_b^{*\top}\mathbf{y}^*}\right)$$
$$= (n-q)\log\kappa - \sum_{i=1}^{n-q}\log(\gamma_i + \kappa) - (n-q)\log\left(1 - \frac{\mathbf{y}^{*\top}\mathbf{U}_b^*\mathbf{D}_b^*\left(\mathbf{D}_b^{*2} + \kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^*\mathbf{U}_b^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}\right).$$

The main part of $2\log B_{10}^{\kappa}$ is

$$T_n^{\kappa} = \frac{\mathbf{y}^{*\top}\mathbf{U}_b^*\mathbf{D}_b^*\left(\mathbf{D}_b^{*2} + \kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^*\mathbf{U}_b^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}.$$

A large value of T_n^{κ} supports the alternative hypothesis. Under the null hypothesis,

$$\mathrm{E}\,T_n^{\kappa} = \frac{1}{n-q}\operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2} + \kappa\mathbf{I}_{n-q})^{-1}\right).$$

Under the alternative hypothesis, consider $\boldsymbol{\beta}_b = c\boldsymbol{\beta}_b^\dagger$ where $\boldsymbol{\beta}_b^\dagger \neq 0$ is a fixed direction and $c > 0$. As $c \to \infty$,

$$T_n^{\kappa} \to \frac{\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^*\mathbf{D}_b^{*2}\left(\mathbf{D}_b^{*2} + \kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*2}\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^\dagger}{\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^*\mathbf{D}_b^{*2}\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^\dagger}.$$

We say T_n^{κ} is consistent along the direction $\boldsymbol{\beta}_b^\dagger$ if

$$\frac{\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^*\mathbf{D}_b^{*2}\left(\mathbf{D}_b^{*2} + \kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*2}\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^\dagger}{\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^*\mathbf{D}_b^{*2}\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^\dagger} > \frac{1}{n-q}\operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2} + \kappa\mathbf{I}_{n-q})^{-1}\right),$$

or equivalently

$$\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^*\left[\mathbf{D}_b^{*2}\left(\mathbf{D}_b^{*2} + \kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*2} - \frac{1}{n-q}\operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2} + \kappa\mathbf{I}_{n-q})^{-1}\right)\mathbf{D}_b^{*2}\right]\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^\dagger > 0.$$

Let k_κ be the number of positive eigenvalues of

$$\mathbf{V}_b^* \left[\mathbf{D}_b^{*2} \left(\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q}\right)^{-1} \mathbf{D}_b^{*2} - \frac{1}{n-q} \operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1}\right) \mathbf{D}_b^{*2}\right] \mathbf{V}_b^{*\top}.$$

Let $\mathcal{S}_\kappa$ be the linear space spanned by the first k_κ columns of $\mathbf{V}_b^*$. Denote by $\mathcal{S}_\kappa^\perp$ the orthogonal complement space of $\mathcal{S}_\kappa$. We have $\mathbb{R}^p = \mathcal{S}_\kappa \oplus \mathcal{S}_\kappa^\perp$. If $\boldsymbol{\beta}_b^\dagger \in \mathcal{S}_\kappa$,

$$\mathbf{V}_b^* \left[\mathbf{D}_b^{*2} \left(\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q}\right)^{-1} \mathbf{D}_b^{*2} - \frac{1}{n-q} \operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1}\right) \mathbf{D}_b^{*2}\right] \mathbf{V}_b^{*\top} > 0.$$

On the other hand, if $\boldsymbol{\beta}_b^\dagger \in \mathcal{S}_\kappa^\perp$,

$$\mathbf{V}_b^* \left[\mathbf{D}_b^{*2} \left(\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q}\right)^{-1} \mathbf{D}_b^{*2} - \frac{1}{n-q} \operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1}\right) \mathbf{D}_b^{*2}\right] \mathbf{V}_b^{*\top} \leq 0.$$

We would like to choose a hyperparameter κ which consists the most consistent directions. To achieve this, we maximize k_κ with respect to κ.

Proposition 2. *For $\kappa_2 > \kappa_1 > 0$, we have $k_{\kappa_1} \geq k_{\kappa_2}$. That is, k_κ ($\kappa > 0$) is decreasing in κ.*

The proposition implies that we should put κ as small as possible. This motivates us to consider $B_{10}^0 = \lim_{\kappa \to 0} B_{10}^\kappa$. It is straightforward to show that

$$2 \log B_{10}^0 = -\sum_{i=1}^{n-q} \log(\gamma_i) + (n-q) \log\left(\frac{\mathbf{y}^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\mathbf{y}^*}\right).$$

B_{10}^0 can be regarded as the Bayes factor with respect to noninformative prior.

Define

$$T_n = \frac{\mathbf{y}^{*\top}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*}.$$

Then we reject the null hypothesis if T_n is small. It can be seen that under the null hypohtesis,

$$T_n \sim \frac{\sum_{i=1}^{n-q} \gamma_i^{-1} Z_i^2}{\sum_{i=1}^{n-q} Z_i^2},$$

where γ_i is the ith eigenvalue of $\mathbf{X}_b^*\mathbf{X}_b^{*\top}$, $i = 1, \ldots, n-q$, and $Z_1, \ldots, Z_{n-q}$ are iid $\mathcal{N}(0,1)$ random variables.

3 Asymptotic results

Let $\boldsymbol{\varepsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$, where ϵ_i's are iid random variable. Denote $\mu_k = \operatorname{E} \epsilon_1^k$. Then $\mu_1 = 0$, $\mu_2 = \phi^{-1}$.

Assumption 1. *Suppose*

Lemma 1. *If $\phi^2 \mu_4 = o(n-q)$,*

$$\mathbf{y}^{*\top}\mathbf{y}^* = (1 + o_P(1))\left(\boldsymbol{\beta}_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b \boldsymbol{\beta}_b + \phi^{-1}(n-q)\right).$$

Proof.

$$\mathbf{y}^{*\top}\mathbf{y}^* = \boldsymbol{\beta}_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b\boldsymbol{\beta}_b + 2\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a)\boldsymbol{\varepsilon}.$$

$$\mathrm{E}\left(\mathbf{y}^{*\top}\mathbf{y}^*\right) = \boldsymbol{\beta}_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b\boldsymbol{\beta}_b + \phi^{-1}(n-q).$$

$$\mathrm{Var}\left(\mathbf{y}^{*\top}\mathbf{y}^*\right) \leq 2\,\mathrm{Var}\left(2\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b\boldsymbol{\beta}_b\right) + 2\,\mathrm{Var}\left(\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a)\boldsymbol{\varepsilon}\right)$$

From (i) of (Chen et al., 2010, Proposition A.1),

$$\mathrm{Var}\left(\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a)\boldsymbol{\varepsilon}\right) = \phi^{-2}\left((\phi^2\mu_4 - 3)\sum_{i=1}^n ((\mathbf{I}_n - \mathbf{P}_a)_{i,i})^2 + 2(n-q)\right) \leq \phi^{-2}(2+\phi^2\mu_4)(n-q).$$

Then

$$\mathrm{Var}\left(\mathbf{y}^{*\top}\mathbf{y}^*\right) \leq 8\phi^{-1}\boldsymbol{\beta}_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b\boldsymbol{\beta}_b + 2\phi^{-2}(2+\phi^2\mu_4)(n-q)$$

Thus, if $\phi^2\mu_4 = o(n-q)$, we have

$$\frac{\mathrm{Var}\left(\mathbf{y}^{*\top}\mathbf{y}^*\right)}{\left(\mathrm{E}\left(\mathbf{y}^{*\top}\mathbf{y}^*\right)\right)^2} \to 0,$$

and consequently $\mathbf{y}^{*\top}\mathbf{y}^* = (1+o_P(1))\,\mathrm{E}(\mathbf{y}^{*\top}\mathbf{y}^*)$.

$\square$

Note that under the normality, $T_n - \mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})/(n-q)$ has zero mean.

Theorem 1. *Suppose the rows of $\mathbf{X}_b$ are iid random vectors with distribution $\mathcal{N}(0,\sigma_b^2\mathbf{I}_p)$. Suppose $p/(n-q) \to c \in (1,+\infty)$. Then*

$$\left(\boldsymbol{\beta}_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b\boldsymbol{\beta}_b + \phi^{-1}(n-q)\right)\left(\frac{\mathbf{y}^{*\top}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\right) \rightsquigarrow \mathcal{N}(0,1).$$

Proof. Note that $\mathbf{X}_b^*\mathbf{X}_b^{*\top} \sim \mathrm{Wishart}(p, \sigma_b^2\mathbf{I}_{n-q})$.

$$\frac{\mathbf{y}^{*\top}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q} = \frac{\phi\mathbf{y}^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)\mathbf{y}^*}{\phi\mathbf{y}^{*\top}\mathbf{y}^*}.$$

We have

$$\begin{aligned}
&\phi\mathbf{y}^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)\mathbf{y}^* \\
=&\phi\boldsymbol{\varepsilon}^\top\tilde{\mathbf{U}}_a\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)\tilde{\mathbf{U}}_a^\top\boldsymbol{\varepsilon} \\
&+ 2\phi\boldsymbol{\varepsilon}^\top\tilde{\mathbf{U}}_a\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)\mathbf{X}_b^*\boldsymbol{\beta}_b \\
&+ \phi\boldsymbol{\beta}_b^\top\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)\mathbf{X}_b^*\boldsymbol{\beta}_b \\
=:&A_1 + A_2 + A_3.
\end{aligned}$$

We have $\mathrm{E}(A_1|\mathbf{X}_b) = \mathrm{E}(A_2|\mathbf{X}_b) = 0$. It is also straightforward to see that

$$\begin{aligned}
\mathrm{Var}(A_1|\mathbf{X}_b) &= 2\,\mathrm{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-2}\right) - 2\frac{1}{n-q}\,\mathrm{tr}^2\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right),\\
\mathrm{Var}(A_2|\mathbf{X}_b) &= 4\phi\boldsymbol{\beta}_b^\top\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\boldsymbol{\beta}_b,\\
\mathrm{Cov}(A_1, A_2|\mathbf{X}_b) &= 0.
\end{aligned}$$

By some theory, we have From (Jiang, 1996, Theorem 5.1),

$$\frac{A_1 + A_2}{\sqrt{\mathrm{Var}(A_1|\mathbf{X}_b) + \mathrm{Var}(A_2|\mathbf{X}_b)}} \rightsquigarrow \mathcal{N}(0,1).$$

From lemma 3,

$$\mathrm{Var}(A_1|\mathbf{X}_b) = (1 + o_P(1))2\sigma_b^{-4}p^{-2}(n-q)\left(\nu_{-2,c} - \nu_{-1,c}^2\right).$$

Now we deal with $\mathrm{Var}(A_2|\mathbf{X}_b)$. Let $\mathbf{O}$ be a $p \times p$ random matrix with Haar distribution which is independent of $\mathbf{X}_b$. The rotation invariance of normal distribution implies that $\mathbf{X}_b\mathbf{O}$ has the same distribution as $\mathbf{X}_b$ and is independent of $\mathbf{O}$. Then

$$\begin{aligned}
\mathrm{Var}(A_2|\mathbf{X}_b) &= 4\phi\boldsymbol{\beta}_b^\top\mathbf{O}\mathbf{O}^\top\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{O}\mathbf{O}^\top\mathbf{X}_b^{*\top})^{-1} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{O}\mathbf{O}^\top\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\mathbf{O}\mathbf{O}^\top\boldsymbol{\beta}_b\\
&\overset{d}{=} 4\phi\boldsymbol{\beta}_b^\top\mathbf{O}\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\mathbf{O}^\top\boldsymbol{\beta}_b.
\end{aligned}$$

Note that $\mathbf{O}^\top\boldsymbol{\beta}_b/\|\mathbf{O}^\top\boldsymbol{\beta}_b\|$ is uniformly distributed on the unit sphere S^{p-1}. From Lemma 2,

$$\begin{aligned}
&\mathrm{E}\left(4\phi\boldsymbol{\beta}_b^\top\mathbf{O}\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\mathbf{O}^\top\boldsymbol{\beta}_b\middle|\mathbf{X}_b\right)\\
=&4\phi p^{-1}\|\boldsymbol{\beta}_b\|^2\,\mathrm{tr}\left(\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\right)\\
=&4\phi p^{-1}\|\boldsymbol{\beta}_b\|^2\left(\frac{1}{(n-q)^2}\,\mathrm{tr}^2\left[(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right]\mathrm{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top}) - \mathrm{tr}\left[(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right]\right).
\end{aligned}$$

Then Lemma 3 implies that

$$\begin{aligned}
&\mathrm{E}\left(4\phi\boldsymbol{\beta}_b^\top\mathbf{O}\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\mathbf{O}^\top\boldsymbol{\beta}_b\middle|\mathbf{X}_b\right)\\
=&(1 + o_P(1))4\phi\|\boldsymbol{\beta}_b\|^2\sigma_b^{-2}p^{-2}(n-q)\left(\nu_{1,c}\nu_{-1,c}^2 - \nu_{-1,c}\right).
\end{aligned}$$

Similarly,

$$
\begin{aligned}
&\mathrm{Var}\left(4\phi\boldsymbol{\beta}_b^\top\mathbf{O}\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}-\frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\mathbf{O}^\top\boldsymbol{\beta}_b\middle|\mathbf{X}_b\right)\\
\leq&\frac{32}{p^2}\phi^2\|\boldsymbol{\beta}_b\|^4\Bigg(\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-2})+\frac{1}{(n-q)^2}\mathrm{tr}^4((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})\,\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^2)\\
&+2\,\mathrm{tr}^2((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})-\frac{4}{(n-q)^2}\mathrm{tr}^3((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})\,\mathrm{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})\Bigg)\\
=&(1+o_P(1))32\phi^2\|\boldsymbol{\beta}_b\|^4\sigma_b^{-4}p^{-4}(n-q)\left(\mathrm{E}(\xi^{-2})+2\left(\mathrm{E}(\xi^{-1})\right)^2-4\left(\mathrm{E}(\xi^{-1})\right)^3\mathrm{E}(\xi)+\left(\mathrm{E}(\xi^{-1})\right)^4\mathrm{E}(\xi^2)\right)\\
=&o_P(1)\left(\mathrm{E}\left(4\boldsymbol{\beta}_b^\top\mathbf{O}\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}-\frac{\mathrm{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\mathbf{O}^\top\boldsymbol{\beta}_b\middle|\mathbf{X}_b\right)\right)^2
\end{aligned}
$$

Thus, we conclude that

$$
\mathrm{Var}(A_2|\mathbf{X}_b)=(1+o_P(1))4\phi\|\boldsymbol{\beta}_b\|^2\sigma_b^{-2}p^{-2}(n-q)\left(\nu_{1,c}\nu_{-1,c}^2-\nu_{-1,c}\right).
$$

Using the same technique, one can show that

$$
A_3=(1+o_P(1))\phi\|\boldsymbol{\beta}_b\|^2p^{-1}(n-q)\left(1-\nu_{1,c}\nu_{-1,c}\right).
$$

From lemma 1, we have

$$
\phi\mathbf{y}^{*\top}\mathbf{y}^*=(1+o_P(1))\left(\phi\boldsymbol{\beta}_b^\top\mathbf{X}_b^{*\top}\mathbf{X}_b^*\boldsymbol{\beta}_b+(n-q)\right).
$$

Using the same technique, we have

$$
\phi\mathbf{y}^{*\top}\mathbf{y}^*=(1+o_P(1))(n-q)\left(1+\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2\nu_{1,c}\right).
$$

We have

$$
\begin{aligned}
&\sqrt{\mathrm{Var}(A_1|\mathbf{X}_b)+\mathrm{Var}(A_2|\mathbf{X}_b)}\\
=&(1+o_P(1))\sqrt{2\sigma_b^{-2}p^{-2}(n-q)\left(\sigma_b^{-2}(\nu_{-2,c}-\nu_{-1,c}^2)+2\phi\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2-\nu_{-1,c})\right)}.
\end{aligned}
$$

Since $[-\infty,0]$ is compact, for every subsequence of $\{n\}$, there is a further subsequence along which

$$
\frac{\phi\|\boldsymbol{\beta}_b\|^2p^{-1}(n-q)\left(1-\nu_{1,c}\nu_{-1,c}\right)}{\sqrt{2\sigma_b^{-2}p^{-2}(n-q)\left(\sigma_b^{-2}(\nu_{-2,c}-\nu_{-1,c}^2)+2\phi\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2-\nu_{-1,c})\right)}}\to s\in[-\infty,0].
$$

Then along this further subsequence, we have

$$
\frac{A_1+A_2+A_3}{\sqrt{2\sigma_b^{-2}p^{-2}(n-q)\left(\sigma_b^{-2}(\nu_{-2,c}-\nu_{-1,c}^2)+2\phi\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2-\nu_{-1,c})\right)}}\rightsquigarrow\mathcal{N}(s,1).
$$

If $s = -\infty$, the above expression means the left hand side tends to $-\infty$ in probability. Thus

$$\left(T_n - \operatorname{tr}((\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1})/(n-q)\right) \frac{(n-q)\left(1 + \phi\sigma_b^2 \|\boldsymbol{\beta}_b\|^2 \nu_{1,c}\right)}{\sqrt{2\sigma_b^{-2} p^{-2}(n-q)\left(\sigma_b^{-2}(\nu_{-2,c} - \nu_{-1,c}^2) + 2\phi\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2 - \nu_{-1,c})\right)}}$$

$$\rightsquigarrow \mathcal{N}(s, 1).$$

Under the null hypothesis,

$$\left(T_n - \operatorname{tr}((\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1})/(n-q)\right) \frac{n-q}{\sqrt{2\sigma_b^{-4} p^{-2}(n-q)(\nu_{-2,c} - \nu_{-1,c}^2)}} \rightsquigarrow \mathcal{N}(0, 1).$$

Thus the critical value of $T_n - \operatorname{tr}((\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1})/(n-q)$ is

$$\Phi(\alpha) \frac{\sqrt{2\sigma_b^{-4} p^{-2}(n-q)(\nu_{-2,c} - \nu_{-1,c}^2)}}{n-q}.$$

Then the power function is

$$\begin{aligned}
&\Pr\left(T_n - \operatorname{tr}((\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1})/(n-q) \le \Phi(\alpha) \frac{\sqrt{2\sigma_b^{-4} p^{-2}(n-q)(\nu_{-2,c} - \nu_{-1,c}^2)}}{n-q}\right) \\
=&\Pr\left(\mathcal{N}(s,1) \le \Phi(\alpha) \frac{(n-q)\left(1 + \phi\sigma_b^2 \|\boldsymbol{\beta}_b\|^2 \nu_{1,c}\right)\sqrt{2\sigma_b^{-4} p^{-2}(n-q)(\nu_{-2,c} - \nu_{-1,c}^2)}}{(n-q)\sqrt{2\sigma_b^{-2} p^{-2}(n-q)\left(\sigma_b^{-2}(\nu_{-2,c} - \nu_{-1,c}^2) + 2\phi\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2 - \nu_{-1,c})\right)}}\right) + o(1) \\
=&\Pr\left(\mathcal{N}(s,1) \le \Phi(\alpha) \frac{\left(1 + \phi\sigma_b^2 \|\boldsymbol{\beta}_b\|^2 \nu_{1,c}\right)}{\sqrt{1 + \frac{2\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2 - \nu_{-1,c})}{\nu_{-2,c} - \nu_{-1,c}^2}}}\right) + o(1) \\
=&\Phi\left(-s + \Phi(\alpha) \frac{\left(1 + \phi\sigma_b^2 \|\boldsymbol{\beta}_b\|^2 \nu_{1,c}\right)}{\sqrt{1 + \frac{2\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2 - \nu_{-1,c})}{\nu_{-2,c} - \nu_{-1,c}^2}}}\right) + o(1) \\
=&\Phi\left(\frac{\phi\|\boldsymbol{\beta}_b\|^2 p^{-1}(n-q)\left(\nu_{1,c}\nu_{-1,c} - 1\right)}{\sqrt{2\sigma_b^{-2} p^{-2}(n-q)\left(\sigma_b^{-2}(\nu_{-2,c} - \nu_{-1,c}^2) + 2\phi\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2 - \nu_{-1,c})\right)}} + \Phi(\alpha) \frac{\left(1 + \phi\sigma_b^2 \|\boldsymbol{\beta}_b\|^2 \nu_{1,c}\right)}{\sqrt{1 + \frac{2\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2 - \nu_{-1,c})}{\nu_{-2,c} - \nu_{-1,c}^2}}}\right) \\
=&\Phi\left(\frac{\sqrt{n-q}\,\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2\left(\nu_{1,c}\nu_{-1,c} - 1\right)}{\sqrt{2\left((\nu_{-2,c} - \nu_{-1,c}^2) + 2\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2 - \nu_{-1,c})\right)}} + \Phi(\alpha) \frac{\left(1 + \phi\sigma_b^2 \|\boldsymbol{\beta}_b\|^2 \nu_{1,c}\right)}{\sqrt{1 + \frac{2\phi\sigma_b^2\|\boldsymbol{\beta}_b\|^2(\nu_{1,c}\nu_{-1,c}^2 - \nu_{-1,c})}{\nu_{-2,c} - \nu_{-1,c}^2}}}\right) + o(1)
\end{aligned}$$

□

Lemma 2. *Let $\mathbf{A}$ be an $p \times p$ symmetric matrix. Let Z be a p dimensional random vector with uniform distribution on the unit sphere S^{p-1}. Then*

$$\mathrm{E}(Z^\top \mathbf{A} Z) = \frac{1}{p}\operatorname{tr}(\mathbf{A}), \quad \operatorname{Var}(Z^\top \mathbf{A} Z) = \frac{2}{p(p+2)}\left(\operatorname{tr}(\mathbf{A}^2) - \frac{1}{p}\operatorname{tr}^2(\mathbf{A})\right) \le \frac{2}{p^2}\operatorname{tr}(\mathbf{A}^2).$$

Proof. The result follows from direct calculation and the fact that for nonnegative integers $k_1, \ldots, k_p$,

$$\mathrm{E} \prod_{i=1}^{p} z_i^{2k_i} = \frac{\Gamma(p/2) \prod_{i=1}^{p} \Gamma(k_i + 1/2)}{\pi^{p/2} \Gamma(\sum_{i=1}^{p} k_i + p/2)},$$

where z_i is the ith coordinate of Z. $\square$

The following lemma is a direct consequence of MP law and Bai Yin law.

Lemma 3. *Under the assumptions of Theorem 1, for every $r \in \mathbb{R}$,*

$$\frac{1}{\sigma_b^{2r} p^r (n-q)} \operatorname{tr}((\mathbf{X}_b^* \mathbf{X}_b^{*\top})^r) \xrightarrow{a.s.} \nu_{r,c}$$

where $\nu_{r,c} = \mathrm{E}\, \xi^r$,, and ξ is a random variable with density function

$$p_c(x) = \mathbf{1}_{\left[(1-c^{-1/2})^2, (1+c^{-1/2})^2\right]}(x) \frac{c}{2\pi x} \sqrt{4/c - (x - (1/c + 1))^2}.$$

Appendices

Appendix A haha1

Proof of Proposition 1. We assume $0 < \alpha < 1$ since the case $\alpha = 0$ or 1 is trivial. Note that the condition implies $\int [\varphi(\mathbf{y}) - \alpha] \mathcal{N}_n(0, \phi^{-1} \mathbf{I}_n)(d\mathbf{y}) = 0$. Hence it suffices to prove $\varphi(\mathbf{y}) \geq \alpha$, a.s. We prove this by contradiction. Suppose $\lambda(\{\mathbf{y} : \varphi(\mathbf{y}) < \alpha\}) > 0$. Then there exists a $\eta > 0$, such that $\lambda(\{\mathbf{y} : \varphi(\mathbf{y}) < \alpha - \eta\}) > 0$. We denote $E = \{\mathbf{y} : \varphi(\mathbf{y}) < \alpha - \eta\}$. From Lebesgue density theorem (Cohn, 2013, Corollary 6.2.6), there exists a point $z \in E$, such that, for each $\epsilon > 0$ there is a $\delta_\epsilon > 0$ such that

$$\left| \frac{\lambda(E^{\complement} \cap C_\epsilon)}{\lambda(C_\epsilon)} \right| < \epsilon,$$

where $C_\epsilon = \prod_{i=1}^{n} [z_i - \delta_\epsilon, z_i + \delta_\epsilon]$. We put

$$\epsilon = \left(\frac{\sqrt{\pi}}{\sqrt{2} \Phi^{-1} \left(1 - \frac{\eta}{6n}\right)} \right)^n \frac{\eta}{3}.$$

Then for any $\phi > 0$,

$$
\begin{aligned}
\alpha \leq & \int_{\mathbb{R}^n} \varphi(\mathbf{y}) \mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) \\
= & \int_{E \cap C_\epsilon} \varphi(\mathbf{y}) \mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) + \int_{E^\complement \cap C_\epsilon} \varphi(\mathbf{y}) \mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) + \int_{C_\epsilon^\complement} \varphi(\mathbf{y}) \mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) \\
\leq & \alpha - \eta + \int_{E^\complement \cap C_\epsilon} \mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) + \int_{C_\epsilon^\complement} \mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) \\
\leq & \alpha - \eta + \left(\frac{\phi}{2\pi}\right)^{n/2} \lambda(E^\complement \cap C_\epsilon) + 2n\left(1 - \Phi(\sqrt{\phi}\delta_\epsilon)\right) \\
\leq & \alpha - \eta + \left(\frac{\phi}{2\pi}\right)^{n/2} \epsilon(2\delta_\epsilon)^n + 2n\left(1 - \Phi(\sqrt{\phi}\delta_\epsilon)\right) \\
= & \alpha - \eta + \left(\frac{\sqrt{\phi}\delta_\epsilon}{\Phi^{-1}\left(1 - \frac{\eta}{6n}\right)}\right)^n \frac{\eta}{3} + 2n\left(1 - \Phi(\sqrt{\phi}\delta_\epsilon)\right).
\end{aligned}
$$

Putting

$$
\phi = \left(\frac{\Phi^{-1}\left(1 - \frac{\eta}{6n}\right)}{\delta_\epsilon}\right)^2
$$

yields the contradiction $\alpha \leq \alpha - (2/3)\eta$. This completes the proof.

□

Proof of Proposition 2. For positive integer m, define $[m] = \{1, , \ldots, m\}$. For a set A, denote by $|A|$ its cardinality. We have

$$
\begin{aligned}
k_\kappa &= \left| \left\{ i \in [n-q] : \frac{\gamma_i^2}{\gamma_i + \kappa} - \frac{1}{n-q} \sum_{j=1}^{n-q} \frac{\gamma_j \gamma_i}{\gamma_j + \kappa} > 0 \right\} \right| \\
&= \left| \left\{ i \in [n-q] : \frac{\gamma_i}{\gamma_i + \kappa} > \frac{1}{n-q} \sum_{j=1}^{n-q} \frac{\gamma_j}{\gamma_j + \kappa} \right\} \right|.
\end{aligned}
$$

Let X be a random variable uniformly distributed on $\{\gamma_1, \ldots, \gamma_{n-q}\}$. That is, $\Pr(X = \gamma_i) = 1/(n-q)$, $i = 1, \ldots, n-q$. Then it can be seen that

$$
k_\kappa = (n-q) \Pr\left(\frac{X}{X+\kappa} > \mathrm{E}\left[\frac{X}{X+\kappa}\right]\right).
$$

Hence we only need to verify

$$
\Pr\left(\frac{X}{X+\kappa_1} > \mathrm{E}\left[\frac{X}{X+\kappa_1}\right]\right) \geq \Pr\left(\frac{X}{X+\kappa_2} > \mathrm{E}\left[\frac{X}{X+\kappa_2}\right]\right). \tag{2}
$$

Let $Y = X/(X + \kappa_2)$. Then

$$
\frac{X}{(X+\kappa_1)} = \frac{\kappa_2 Y}{\kappa_1 + (\kappa_2 - \kappa_1)Y} := f(Y).
$$

Note that $f(Y)$ is increasing for $Y \geq 0$. Then the inequality (2) is equivalent to

$$\Pr\left(Y > f^{-1}\left(\mathrm{E}\, f(Y)\right)\right) \geq \Pr\left(Y > \mathrm{E}\, Y\right).$$

Hence we only need to verify $f^{-1}\left(\mathrm{E}\, f(Y)\right) \leq \mathrm{E}\, Y$, or equivalently, $\mathrm{E}\, f(Y) \leq f(\mathrm{E}\, Y)$. But the last inequality is a direct consequence of the concavity of $f(Y)$. This completes the proof.

□

Appendix B haha3

Theorem 2. *Let* $\mathbf{A}$ *be an* $n \times n$ *symmetric matrix. Suppose the elements of* $\boldsymbol{\varepsilon}$ *are symmetric and have finite eighth moments. Then*

$$\frac{\boldsymbol{\varepsilon}^\top \mathbf{A} \boldsymbol{\varepsilon} - \sigma^2 \operatorname{tr}(\mathbf{A})}{\sqrt{2\sigma^4 \operatorname{tr}(\mathbf{A}^2) + (\mu_4 - 3\sigma^4) \operatorname{tr}(\mathbf{A} \circ \mathbf{A})}}$$

Proof. Let

$$\tilde{a}_{i,j} = \frac{a_{i,j}}{\sqrt{2\sigma^4 \operatorname{tr}(\mathbf{A}^2) + (\mu_4 - 3\sigma^4) \operatorname{tr}(\mathbf{A} \circ \mathbf{A})}}$$

Then

$$\frac{\boldsymbol{\varepsilon}^\top \mathbf{A} \boldsymbol{\varepsilon} - \sigma^2 \operatorname{tr}(\mathbf{A})}{\sqrt{2\sigma^4 \operatorname{tr}(\mathbf{A}^2) + (\mu_4 - 3\sigma^4) \operatorname{tr}(\mathbf{A} \circ \mathbf{A})}} = \sum_{i=1}^n \tilde{a}_{i,i} \epsilon_i^2 + \sum_{i=1}^n \sum_{j \neq i}^n \tilde{a}_{i,j} \epsilon_i \epsilon_j$$

Let $z_1, \ldots, z_n$ be iid random variables with distribution $\mathcal{N}(0, \tilde{\sigma}^2)$ and $\check{z}_1, \ldots, \check{z}_n$ be iid random variables with distribution $\mathcal{N}(0, \tilde{\mu}_4 - \tilde{\sigma}^4)$ which are independent of $\epsilon_1, \ldots, \epsilon_n$. For $l = 1, \ldots, n$, let

$$\begin{aligned}
S_l =& \sum_{i=1}^{l-1} \tilde{a}_{i,i}(\epsilon_i^2 - \sigma^2) + \sum_{i=l+1}^{n} \tilde{a}_{i,i} \check{z}_i + \sum_{i=1}^{l-1} \sum_{j \neq i}^{l-1} \tilde{a}_{i,j} \epsilon_i \epsilon_j + 2 \sum_{i=1}^{l-1} \sum_{j=l+1}^{n} \tilde{a}_{i,j} \epsilon_i z_j + \sum_{i=l+1}^{n} \sum_{j=l+1}^{n} \tilde{a}_{i,j} z_i z_j \\
\xi_l =& \tilde{a}_{l,l}(\epsilon_l^2 - \sigma^2) + 2 \sum_{i<l} \tilde{a}_{i,l} \epsilon_i \epsilon_l + 2 \sum_{i>l} \tilde{a}_{i,l} z_i \epsilon_l \\
\eta_l =& \tilde{a}_{l,l} \check{z}_l + 2 \sum_{i<l} \tilde{a}_{i,l} \epsilon_i z_l + 2 \sum_{i>l} \tilde{a}_{i,l} z_i z_l
\end{aligned}$$

Then

$$S_n + \xi_n = \sum_{i=1}^n \tilde{a}_{i,i}(\epsilon_i^2 - \sigma^2) + \sum_{i=1}^n \sum_{j \neq i}^n \tilde{a}_{i,j} \epsilon_i \epsilon_j$$

$$S_1 + \eta_1 = \sum_{i=1}^n \tilde{a}_{i,i} \check{z}_i + \sum_{i=1}^n \sum_{j \neq i}^n \tilde{a}_{i,j} z_i z_j$$

For $l = 1, \ldots, n-1$,

$$S_l + \xi_l = S_{l+1} + \eta_{l+1}$$

Thus, for any $f \in \mathscr{C}^4(\mathbb{R})$,

$$\begin{aligned}
&\left| \mathrm{E}\, f\left(\sum_{i=1}^n \tilde{a}_{i,i}(\epsilon_i^2 - \sigma^2) + \sum_{i=1}^n \sum_{j\neq i}^n \tilde{a}_{i,j}\epsilon_i\epsilon_j \right) - \mathrm{E}\, f\left(\sum_{i=1}^n \tilde{a}_{i,i}\check{z}_i + \sum_{i=1}^n \sum_{j\neq i}^n \tilde{a}_{i,j} z_i z_j \right) \right| \\
=& \left| \mathrm{E}\, f(S_n + \xi_n) - \mathrm{E}\, f(S_1 + \eta_1) \right| \\
=& \left| \sum_{l=2}^n \left(\mathrm{E}\, f(S_l + \xi_l) - \mathrm{E}\, f(S_{l-1} + \xi_{l-1}) \right) + \mathrm{E}\, f(S_1 + \xi_1) - \mathrm{E}\, f(S_1 + \eta_1) \right| \\
=& \left| \sum_{l=1}^n \mathrm{E}\, f(S_l + \xi_l) - \mathrm{E}\, f(S_l + \eta_l) \right|
\end{aligned}$$

Apply Taylor's theorem, for $l = 1, \ldots, n$,

$$\begin{aligned}
f(S_l + \xi_l) =& f(S_l) + \xi_l f'(S_l) + \frac{1}{2}\xi_l^2 f''(S_l) + \frac{1}{6}\xi_l^3 f'''(S_l) + \frac{1}{24}\xi_l^4 f''''(S_l + \theta_1\xi_l), \\
f(S_l + \eta_l) =& f(S_l) + \eta_l f'(S_l) + \frac{1}{2}\eta_l^2 f''(S_l) + \frac{1}{6}\eta_l^3 f'''(S_l) + \frac{1}{24}\eta_l^4 f''''(S_l + \theta_2\eta_l),
\end{aligned}$$

where $\theta_1, \theta_2 \in [0, 1]$. Thus,

$$\begin{aligned}
& \left| \mathrm{E}\, f(S_l + \xi_l) - \mathrm{E}\, f(S_l + \eta_l) \right| \\
\leq & \left| \mathrm{E}(\xi_l - \eta_l) f'(S_l) + \mathrm{E}\, \frac{1}{2}(\xi_l^2 - \eta_l^2) f''(S_l) + \mathrm{E}\, \frac{1}{6}(\xi_l^3 - \eta_l^3) f'''(S_l) \right| + \frac{1}{12}\| f'''' \|_\infty \left(\mathrm{E}(\xi_l^4) + \mathrm{E}(\eta_l^4) \right) \\
= & \left| \mathrm{E}\, f'(S_l)\, \mathrm{E}_l(\xi_l - \eta_l) + \mathrm{E}\, \frac{1}{2} f''(S_l)\, \mathrm{E}_l(\xi_l^2 - \eta_l^2) + \mathrm{E}\, \frac{1}{6} f'''(S_l)\, \mathrm{E}_l(\xi_l^3 - \eta_l^3) \right| + \frac{1}{12}\| f'''' \|_\infty \left(\mathrm{E}(\xi_l^4) + \mathrm{E}(\eta_l^4) \right),
\end{aligned}$$

where E_l is the expectation with respect to $\epsilon_l, z_l, \check{z}_l$. It is straightforward to show that

$$\begin{aligned}
& \mathrm{E}_l(\xi_l - \eta_l) = 0, \\
& \mathrm{E}_l(\xi_l^2 - \eta_l^2) = \left((\mu_4 - \sigma^4) - (\tilde{\mu}_4 - \tilde{\sigma}^4) \right) \tilde{a}_{l,l}^2 + 4(\sigma^2 - \tilde{\sigma}^2) \left(\sum_{i<l} \tilde{a}_{i,l}\epsilon_i + \sum_{i>l} \tilde{a}_{i,l} z_i \right)^2, \\
& \mathrm{E}_l(\xi_l^3 - \eta_l^3) = (\mu_6 - 3\mu_4\sigma^2 + 2\sigma^6)\tilde{a}_{l,l}^3 + 12(\mu_4 - \sigma^4)\tilde{a}_{l,l} \left(\sum_{i<l} \tilde{a}_{i,l}\epsilon_i + \sum_{i>l} \tilde{a}_{i,l} z_i \right)^2.
\end{aligned}$$

Thus,

$$
\begin{aligned}
&|\mathrm{E}\, f(S_l+\xi_l) - \mathrm{E}\, f(S_l+\eta_l)| \\
\leq & \frac{1}{2}\|f''\|_\infty \left(\left|(\mu_4-\sigma^4)-(\tilde{\mu}_4-\tilde{\sigma}^4)\right| \tilde{a}_{l,l}^2 + 4|\sigma^2-\tilde{\sigma}^2|\, \mathrm{E} \left(\sum_{i<l} \tilde{a}_{i,l}\epsilon_i + \sum_{i>l} \tilde{a}_{i,l} z_i \right)^2 \right) \\
& + \frac{1}{6}\|f'''\|_\infty \left((\mu_6 - 3\mu_4\sigma^2 + 2\sigma^6)|\tilde{a}_{l,l}^3| + 12(\mu_4-\sigma^4)\, \mathrm{E}\, |\tilde{a}_{l,l}| \left(\sum_{i<l} \tilde{a}_{i,l}\epsilon_i + \sum_{i>l} \tilde{a}_{i,l} z_i \right)^2 \right) \\
& + \frac{1}{12}\|f''''\|_\infty \left(\mathrm{E}(\xi_l^4) + \mathrm{E}(\eta_l^4) \right) \\
= & \frac{1}{2}\|f''\|_\infty \left(\left|(\mu_4-\sigma^4)-(\tilde{\mu}_4-\tilde{\sigma}^4)\right| \tilde{a}_{l,l}^2 + 4|\sigma^2-\tilde{\sigma}^2| \left(\sum_{i<l} \tilde{a}_{i,l}^2\sigma^2 + \sum_{i>l} \tilde{a}_{i,l}^2 \tilde{\sigma}^2 \right) \right) \\
& + \frac{1}{6}\|f'''\|_\infty \left((\mu_6 - 3\mu_4\sigma^2 + 2\sigma^6)|\tilde{a}_{l,l}^3| + 12(\mu_4-\sigma^4)|\tilde{a}_{l,l}| \left(\sum_{i<l} \tilde{a}_{i,l}^2\sigma^2 + \sum_{i>l} \tilde{a}_{i,l}^2 \tilde{\sigma}^2 \right) \right) \\
& + \frac{1}{12}\|f''''\|_\infty \left(\mathrm{E}(\xi_l^4) + \mathrm{E}(\eta_l^4) \right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\sum_{l=1}^n |\mathrm{E}\, f(S_l+\xi_l) - \mathrm{E}\, f(S_l+\eta_l)| \\
\leq & \frac{1}{2}\|f''\|_\infty \left(|\mu_4-\tilde{\mu}_4| \sum_{l=1}^n \tilde{a}_{l,l}^2 + 2|\sigma^4-\tilde{\sigma}^4| \sum_{i=1}^n \sum_{l=1}^n \tilde{a}_{i,l}^2 \right) \\
& + \frac{1}{6}\|f'''\|_\infty \left((\mu_6 - 3\mu_4\sigma^2 + 2\sigma^6) \sum_{l=1}^n |\tilde{a}_{l,l}^3| + 12(\mu_4-\sigma^4) \sum_{l=1}^n |\tilde{a}_{l,l}| \left(\sum_{i<l} \tilde{a}_{i,l}^2\sigma^2 + \sum_{i>l} \tilde{a}_{i,l}^2 \tilde{\sigma}^2 \right) \right) \\
& + \frac{1}{12}\|f''''\|_\infty \left(\mathrm{E}(\xi_l^4) + \mathrm{E}(\eta_l^4) \right).
\end{aligned}
$$

We have

$$
\mathrm{E}(\xi_l^4) = O((\sum_{i=1}^n \tilde{a}_{i,l}^2)^2), \quad \mathrm{E}(\eta_l^4) = O((\sum_{i=1}^n \tilde{a}_{i,l}^2)^2).
$$

Then

$$
\begin{aligned}
&\sum_{l=1}^n |\mathrm{E}\, f(S_l+\xi_l) - \mathrm{E}\, f(S_l+\eta_l)| \\
\leq & \frac{1}{2}\|f''\|_\infty \left(|\mu_4-\tilde{\mu}_4| \sum_{l=1}^n \tilde{a}_{l,l}^2 + 2|\sigma^4-\tilde{\sigma}^4| \sum_{i=1}^n \sum_{l=1}^n \tilde{a}_{i,l}^2 \right) \\
& + O(1)\|f'''\|_\infty \max_{l\in\{1,\ldots,n\}} |\tilde{a}_{l,l}| + O(1)\|f''''\|_\infty \max_{l\in\{1,\ldots,n\}} \left(\sum_{i=1}^n \tilde{a}_{i,l}^2 \right).
\end{aligned}
$$

$\square$

References

Buhlmann, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli*, 19(4):1212–1242.

Chen, S. X., Zhang, L., and Zhong, P. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819.

Cohn, D. L. (2013). *Measure Theory*. Birkhauser Advanced Texts Basler Lehrbucher. Birkhuser Basel, 2 edition.

Feng, L., Zou, C., Wang, Z., and Chen, B. (2013). Rank-based score tests for high-dimensional regression coefficients. *Electronic Journal of Statistics*, 7(0):2131–2149.

Goddard, S. D. and Johnson, V. E. (2016). Restricted most powerful bayesian tests for linear models. *Scandinavian Journal of Statistics*, 43(4):1162–1177.

Goeman, J. J., van de Geer, S. A., and van Houwelingen, H. C. (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):477–493.

Goeman, J. J., van Houwelingen, H. C., and Finos, L. (2011). Testing against a high-dimensional alternative in the generalized linear model: asymptotic type i error control. *Biometrika*, 98(2):381–390.

JAVIER GIRÓN, F., LINA MARTNEZ, M., MORENO, E., and TORRES, F. (2006). Objective testing procedures in linear models: Calibration of the p-values. *Scandinavian Journal of Statistics*, 33(4):765–784.

Jiang, J. (1996). Reml estimation: asymptotic behavior and related topics. *Annals of Statistics*, 24(1):255–286.

Kass, R. E. and Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association*, 90(431):928–934.

Lan, W., Wang, H., and Tsai, C. L. (2014). Testing covariates in high-dimensional regression. *Annals of the Institute of Statistical Mathematics*, 66(2):279–301.

Lan, W., Zhong, P.-S., Li, R., Wang, H., and Tsai, C.-L. (2016). Testing a single regression coefficient in high dimensional linear models. *Journal of Econometrics*, 195(1):154–168.

Maruyama, Y. and George, E. I. (2011). Fully bayes factors with a generalized g -prior. *Ann. Statist.*, 39(5):2740–2765.

Shang, Z. and Clayton, M. K. (2011). Consistency of bayesian linear model selection with a growing number of parameters. *Journal of Statistical Planning and Inference*, 141(11):3463–3474.

Wang, S. and Cui, H. (2015). A new test for part of high dimensional regression coefficients. *Journal of Multivariate Analysis*, 137:187–203.

Xu, K. (2016). A new nonparametric test for high-dimensional regression coefficients. *Journal of Statistical Computation and Simulation*, 87(5):855–869.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242.

Zhong, P. S. and Chen, S. X. (2011). Tests for high-dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association*, 106(493):260–274.

Zhou, Q. and Guan, Y. (2018). On the null distribution of bayes factors in linear regression. *Journal of the American Statistical Association*, 0(0):1–10.